MSR222

What Prompt Engineering Missed and How Context Engineering Fixes It

Hanan Irfan, MSc, ***Tushar Srivastava, MSc***
ConnectHEOR, London, UK
Email: tushar.srivastava@connectheor.com

BACKGROUND

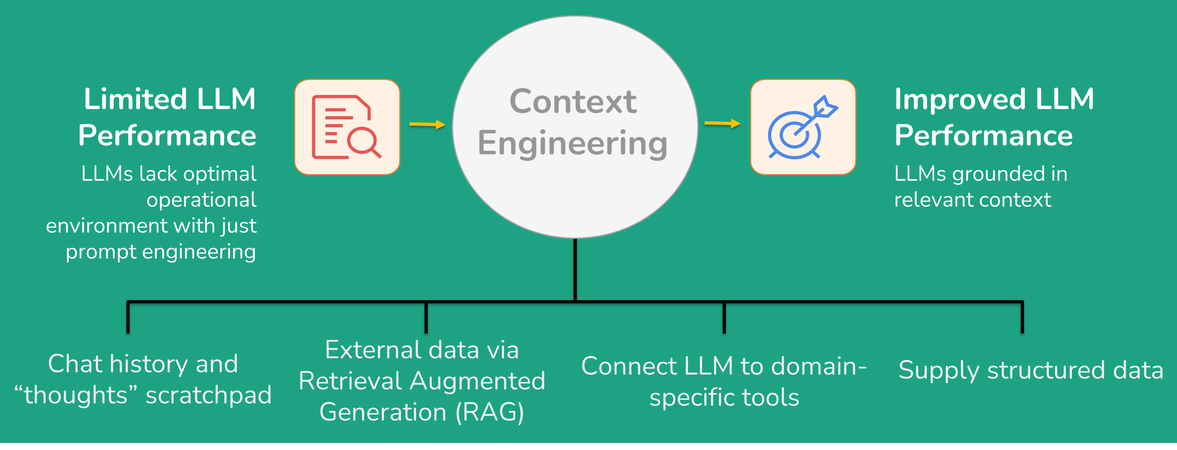
- The use of **large language models (LLMs)** in **health economics and outcomes research (HEOR)** is rapidly increasing, with applications ranging from model explanation, automated reporting to literature review and health technology assessment (HTA) landscape analysis.
- Traditionally, success was thought to depend on **"prompt engineering"**, designing clever instructions or examples to elicit better responses.
- However, in **complex, technical, and niche domains** like HEOR, **prompt engineering alone has proven insufficient**.

GAP: LLMs are often not pretrained on HEOR-specific concepts, methods, or terminology, making them prone to hallucinations and errors when prompts are the only control mechanism. This creates a risk for adoption in real-world HEOR workflows.

OBJECTIVES

This study **compares prompt-only vs. context-engineered AI systems** across multiple HEOR use cases, highlighting the value of structured environments for domain-specific applications.

CONCEPT



METHODOLOGY

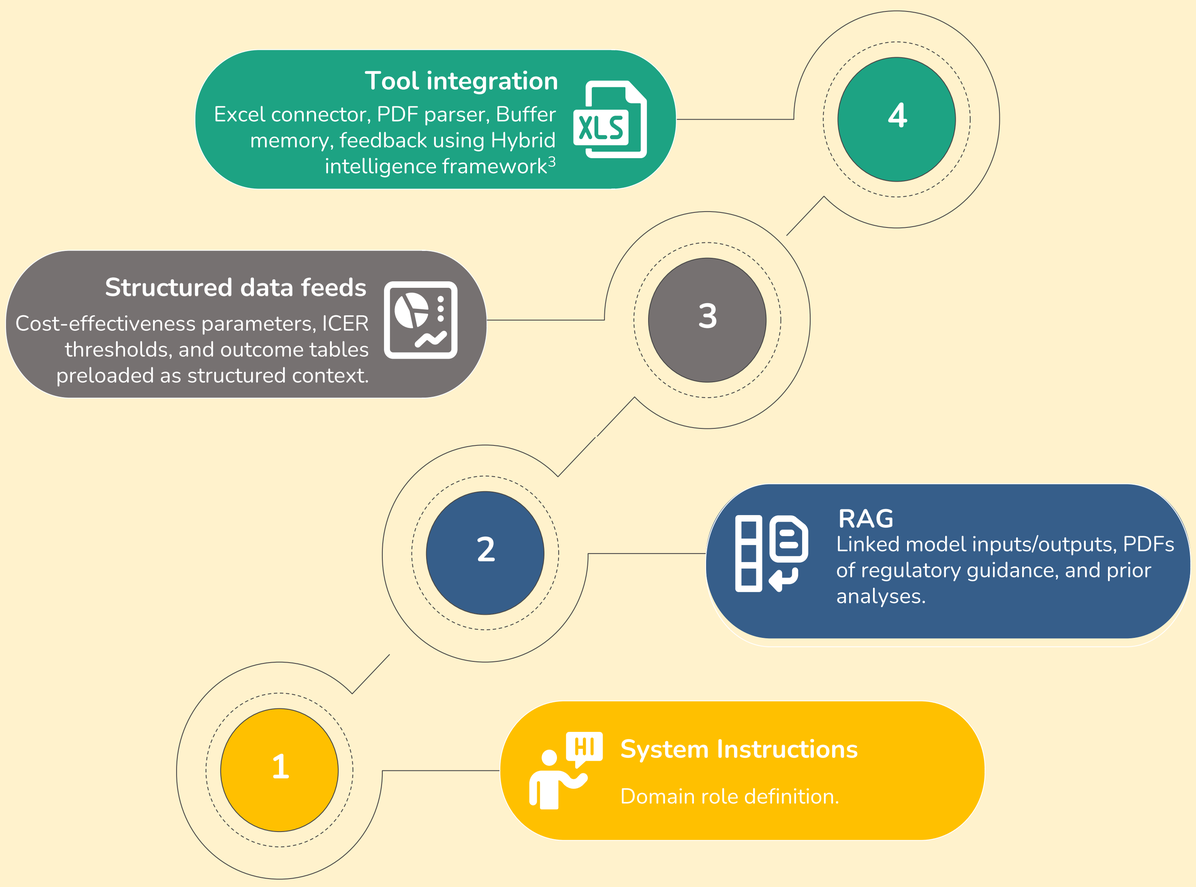
Comparative analysis of LLM performance in **three HEOR tasks**:

- Automated cost-effectiveness report generation
- Chatbot-based model explainer
- Abstract summarization of regulatory documents

Configurations tested:

- **Prompt-only** (single well-crafted prompts with Chain-of-Thought prompting).
- **Context-engineered** (multi-layered context environment).

Context Engineering Implementation



RESULTS

Cost-effectiveness Report Generation

Prompt-only	Context-engineered
Used non-technical phrasing such as "the drug is cost-effective enough," lacking HEOR precision.	Applied formal HTA-compliant terminology: "At a willingness-to-pay threshold of £30,000/QALY, Drug A has a 72% probability of being cost-effective."
Omitted scenario analyses for alternative time horizons, leading to incomplete HTA-ready summaries.	Retrieved all defined scenarios from Excel tabs ("Scenario_10yr")and auto-inserted ICERs in report sections.
Generated base-case summaries inconsistent with model parameters (e.g., reported an ICER of £35,000/QALY while Excel model showed £42,180/QALY for Drug A vs SoC).	Extracted ICER, incremental costs (£12,560), and QALYs (0.298) directly from linked Excel sheets using named cell references, ensuring numeric fidelity.

Chatbot-based Model Explainer

Prompt-only	Context-engineered
Could not clarify disease-specific inputs when prompted ("Which transition probabilities drive the model?").	Referenced data dictionary: "Transition from PFS to PD is 0.052 per month, derived from parametric PFS and OS fits.
Inconsistently responded to sequential queries, e.g., changed reported number of health states between turns.	Maintained state memory, referencing the same model schema across turns (verified identical 3-state structure).
Provided vague text like "This is a Markov model with transitions between health states," without specifying structure.	Retrieved the model JSON/YAML schema to explain: "This is a three-state Markov model (Progression-Free, Progressed Disease, Death) with monthly cycles and a 20-year horizon."

Abstract Summarisation of Regulatory Documents

Prompt-only	Context-engineered
Missed regulatory nuances like "clinical benefit not demonstrated in subgroup ≥ 75 years."	Preserved phrasing: "Evidence insufficient to establish benefit in the ≥ 75 years subgroup," aligning with EMA appraisal wording.
Invented or conflated outcomes not in the PDF (e.g., stated "statistically significant QALY gain" though no QALY data were reported).	Cross-checked extracted data from MHRA submission PDF sections (Section 4.2, Table 12), ensuring only referenced endpoints were summarised.

CONCLUSION

- **Prompt engineering alone is insufficient** for specialized HEOR tasks.
- **Context engineering consistently improved performance.**[4]
- By integrating **RAG, structured data feeds, and domain-specific tools**, outputs were more **aligned with technical expectations**, reducing hallucination and improving consistency.
- **Implication for practice:** As AI adoption in HEOR accelerates, success will hinge not on clever prompts but on **well-structured, context-aware, and tool-integrated environments**.
- **Future direction:** Establishing **best practices and implementation frameworks** for context engineering will be critical to scale LLM deployment across HEOR research and decision-making.

> "Context engineering is essential for adapting LLMs to niche domains like HEOR, enabling accurate, domain-grounded outputs that prompt engineering alone cannot achieve."

REFERENCES

1. Mesko, B., 2025
2. Mei, L., 2025
3. Srivastava, T., 2025
4. Arefeen, A., 2024

Financial Disclosure: Authors are employees of ConnectHEOR Limited. No external funding received to conduct this research and no conflict of interest to declare.

Poster presented at ISPOR in Glasgow, Scotland, 9-12 November 2025